1. Title of your project and group information

Title: Yelp Restaurant Photo Classification

Group Information: Yu Wen, Huijing Zhang, Qian Shen, Shiyu Zhang

6. Evaluation Plan

How will you evaluate your solution?

As we got data from Kaggle website, they offer a test data set for us, so the purpose is to train our training data set and set labels to the test data set, and the result can be examined by accuracy due to Mean F1-Score to exam the accuracy of our data.

Meanwhile, we will also use cross-validation method to see the accuracy of our training data itself.

What evaluation measures are you planning to use.

We will use Mean F1-Score also know as example-based F-measure in the multi-label learning literature. It considers both the [precision](https://en.wikipedia.org/wiki/Precision_(information_retrieval)) *p* and the [recall](https://en.wikipedia.org/wiki/Recall_(information_retrieval)) *r* of the test to compute the score: *p* is the number of correct positive results divided by the number of all positive results, and *r* is the number of correct positive results divided by the number of positive results that should have been returned. The F1 score can be interpreted as a weighted average of the [precision and recall](https://en.wikipedia.org/wiki/Precision_and_recall), where an F1 score reaches its best value at 1 and worst at 0.

The traditional F-measure or balanced F-score (**F1 score**) is the [harmonic mean](https://en.wikipedia.org/wiki/Harmonic_mean#Harmonic_mean_of_two_numbers) of precision and recall:

.

The F1 metric weights recall and precision equally, and a good retrieval algorithm will maximize both precision and recall simultaneously. Thus, moderately good performance on both will be favored over extremely good performance on one and poor performance on the other.

Also, when we need to do classify, we think we can also draw an [ROC curve](http://stats.stackexchange.com/questions/7207/roc-vs-precision-and-recall-curves/7210) during we are choosing the threshold of our algorithm. ROC analysis provides tools to select possibly optimal models and to discard suboptimal ones independently from (and prior to specifying) the cost context or the class distribution.